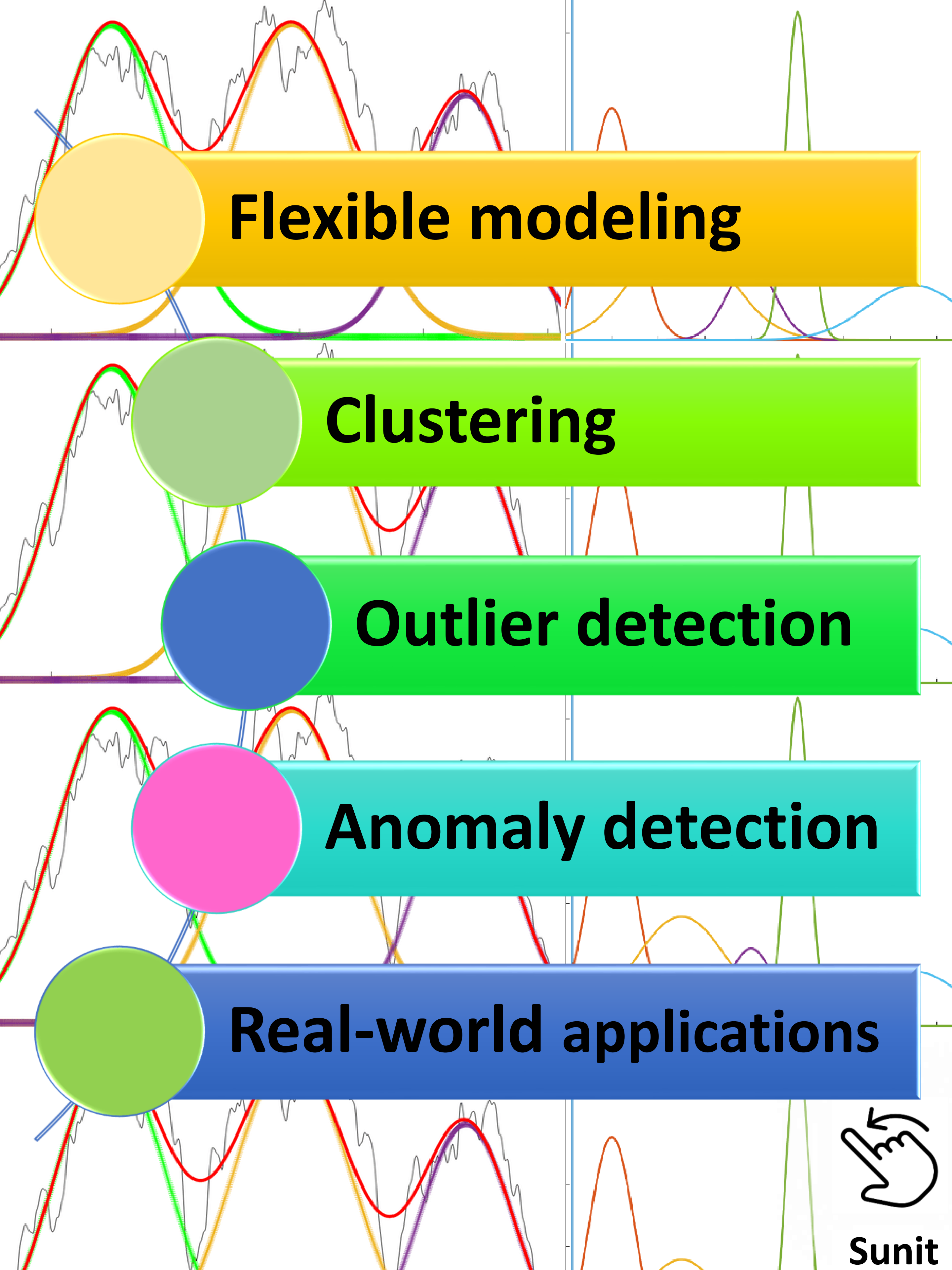
Gaussian Mixture Models

Sunit



Flexible modeling

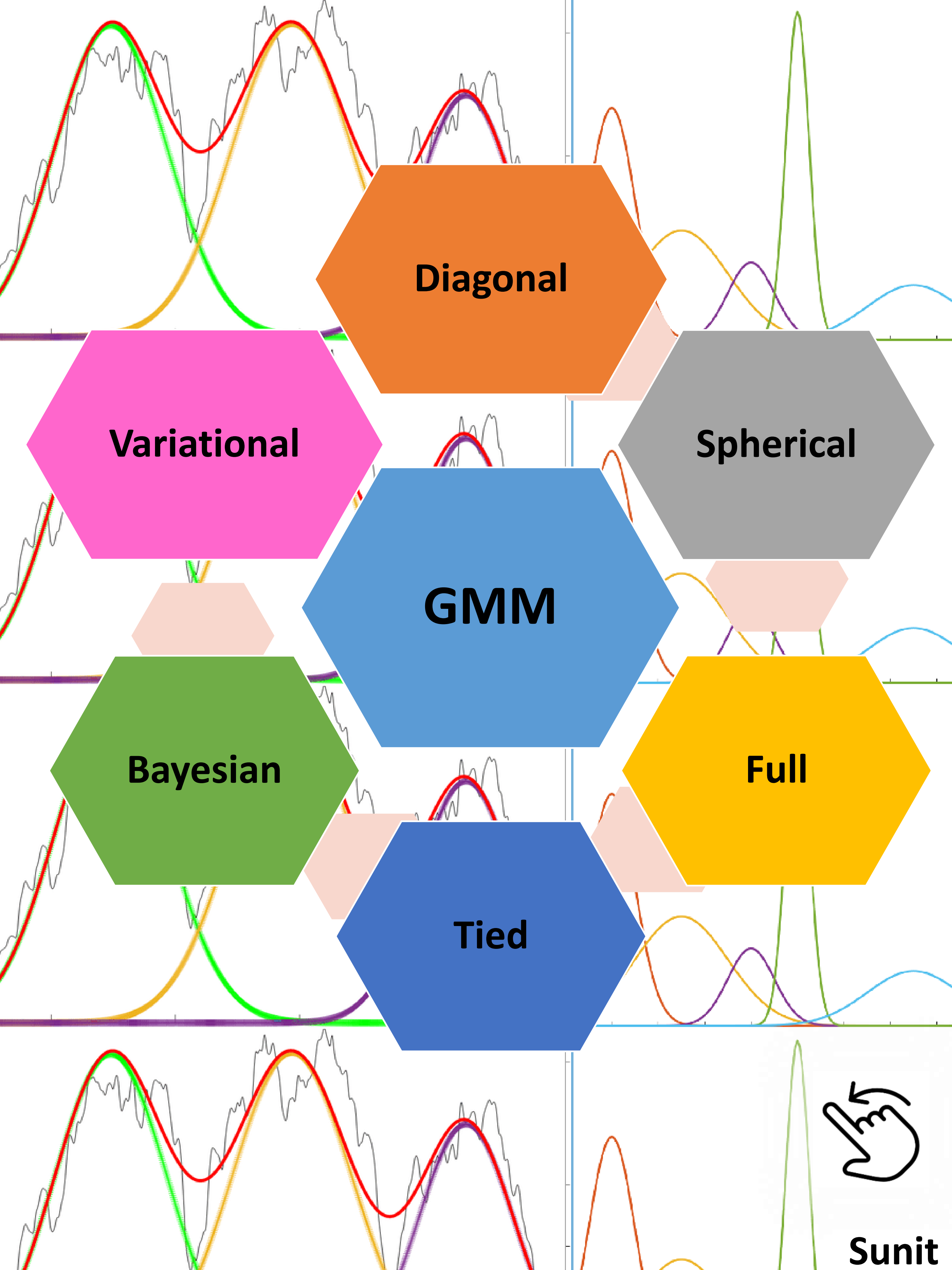
Clustering

Outlier detection

Anomaly detection

Real-world applications

Sunit



Diagonal

Variational

Spherical

GMM

Bayesian

Full

Tied

Sunit

Flexible	• Combines multiple Gaussian (normal) distributions
Clustering	• Useful when there is no prior knowledge of the number of clusters or when the clusters have different shapes and sizes
Outlier Detection	• Models the data distribution and then identifies the data points that are far from the distribution
Anomaly Detection	• Models the normal data distribution and then identifies data points that are unlikely to have come from that distribution
Real-world applications	• Used in computer vision to segment images into different regions based on color or texture similarity

Sunit

Flexible

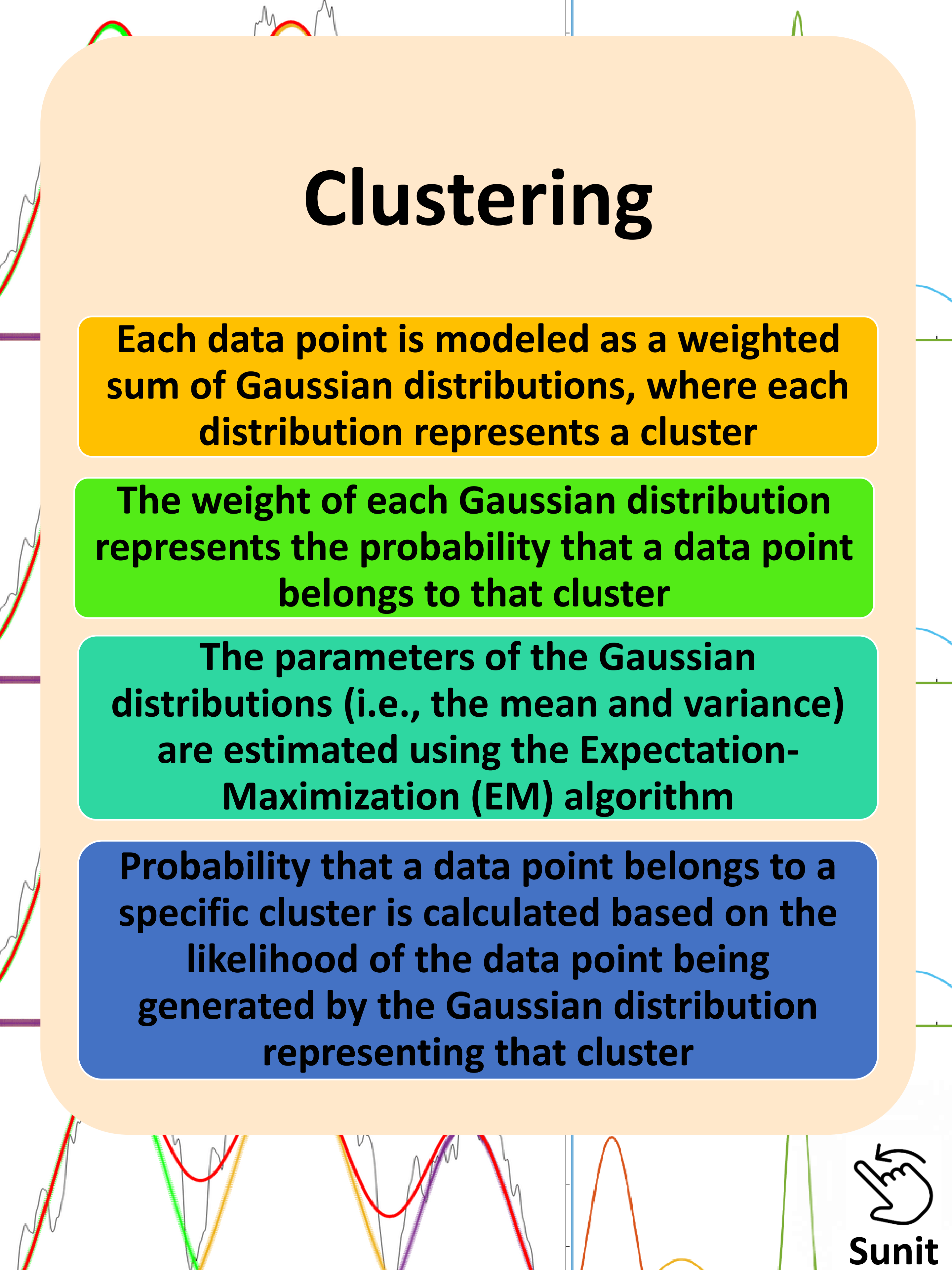
Represents complex data distributions using a combination of simpler Gaussian (normal) distributions

Represents a dataset as a weighted sum of Gaussian distributions, where each distribution has its own mean and variance

Combination of multiple Gaussian distributions allows the GMM to capture a variety of patterns and structures in the data

If the data has many distinct peaks, we can increase the number of Gaussian components to capture the structure of the data more accurately

Sunit



Clustering

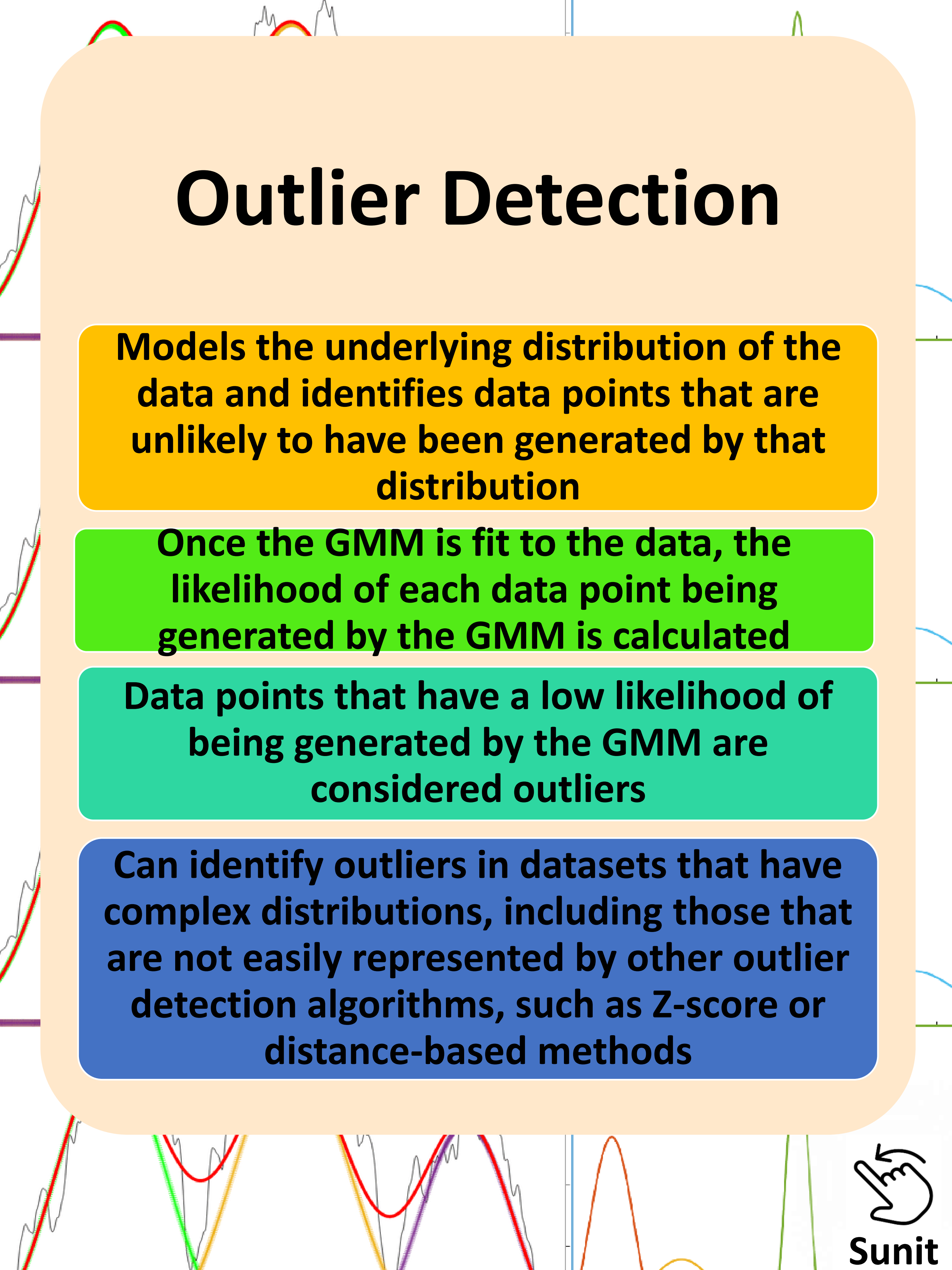
Each data point is modeled as a weighted sum of Gaussian distributions, where each distribution represents a cluster

The weight of each Gaussian distribution represents the probability that a data point belongs to that cluster

The parameters of the Gaussian distributions (i.e., the mean and variance) are estimated using the Expectation-Maximization (EM) algorithm

Probability that a data point belongs to a specific cluster is calculated based on the likelihood of the data point being generated by the Gaussian distribution representing that cluster

Sunit



Outlier Detection

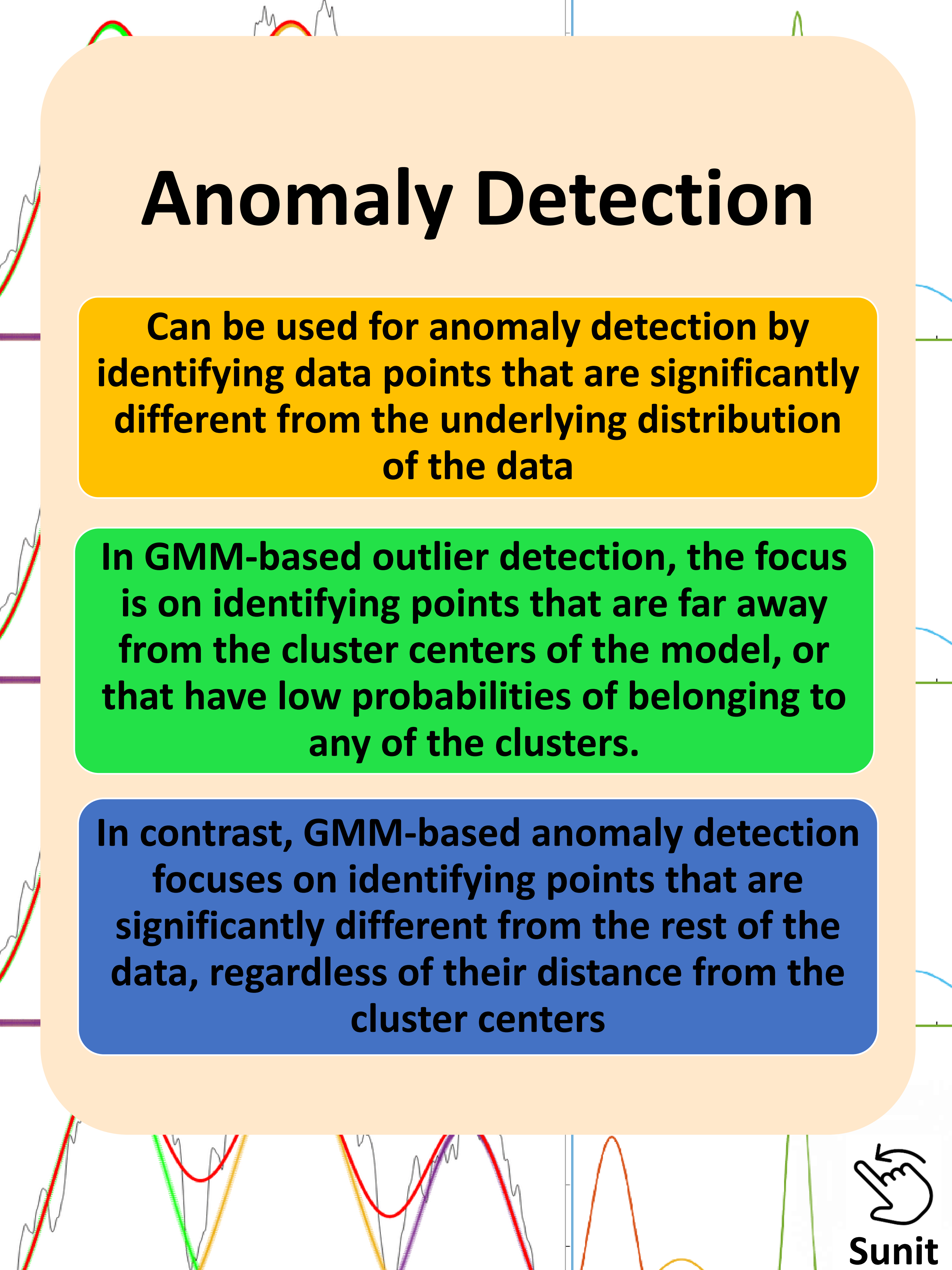
Models the underlying distribution of the data and identifies data points that are unlikely to have been generated by that distribution

Once the GMM is fit to the data, the likelihood of each data point being generated by the GMM is calculated

Data points that have a low likelihood of being generated by the GMM are considered outliers

Can identify outliers in datasets that have complex distributions, including those that are not easily represented by other outlier detection algorithms, such as Z-score or distance-based methods

Sunit



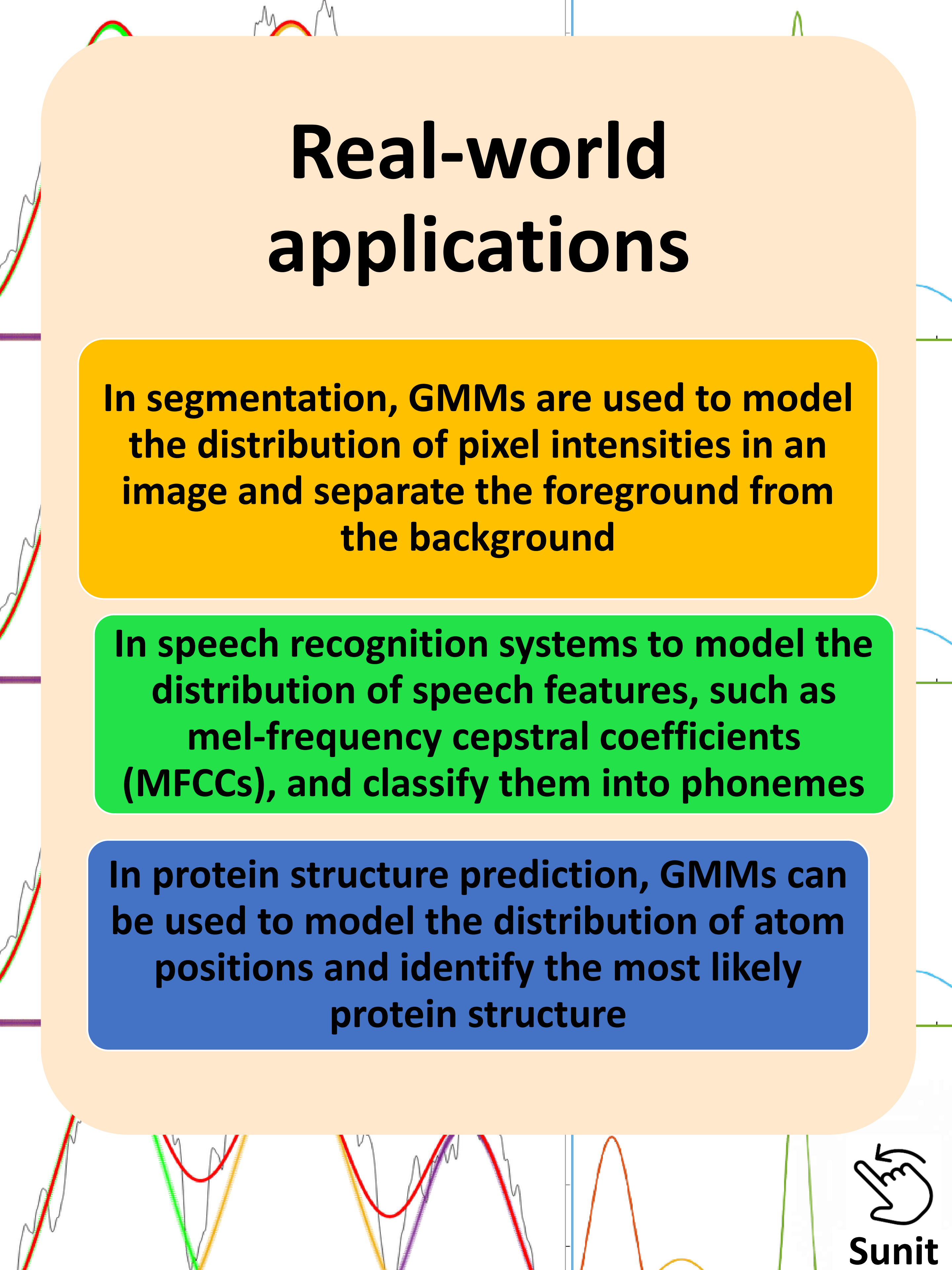
Anomaly Detection

Can be used for anomaly detection by identifying data points that are significantly different from the underlying distribution of the data

In GMM-based outlier detection, the focus is on identifying points that are far away from the cluster centers of the model, or that have low probabilities of belonging to any of the clusters.

In contrast, GMM-based anomaly detection focuses on identifying points that are significantly different from the rest of the data, regardless of their distance from the cluster centers

Sunit



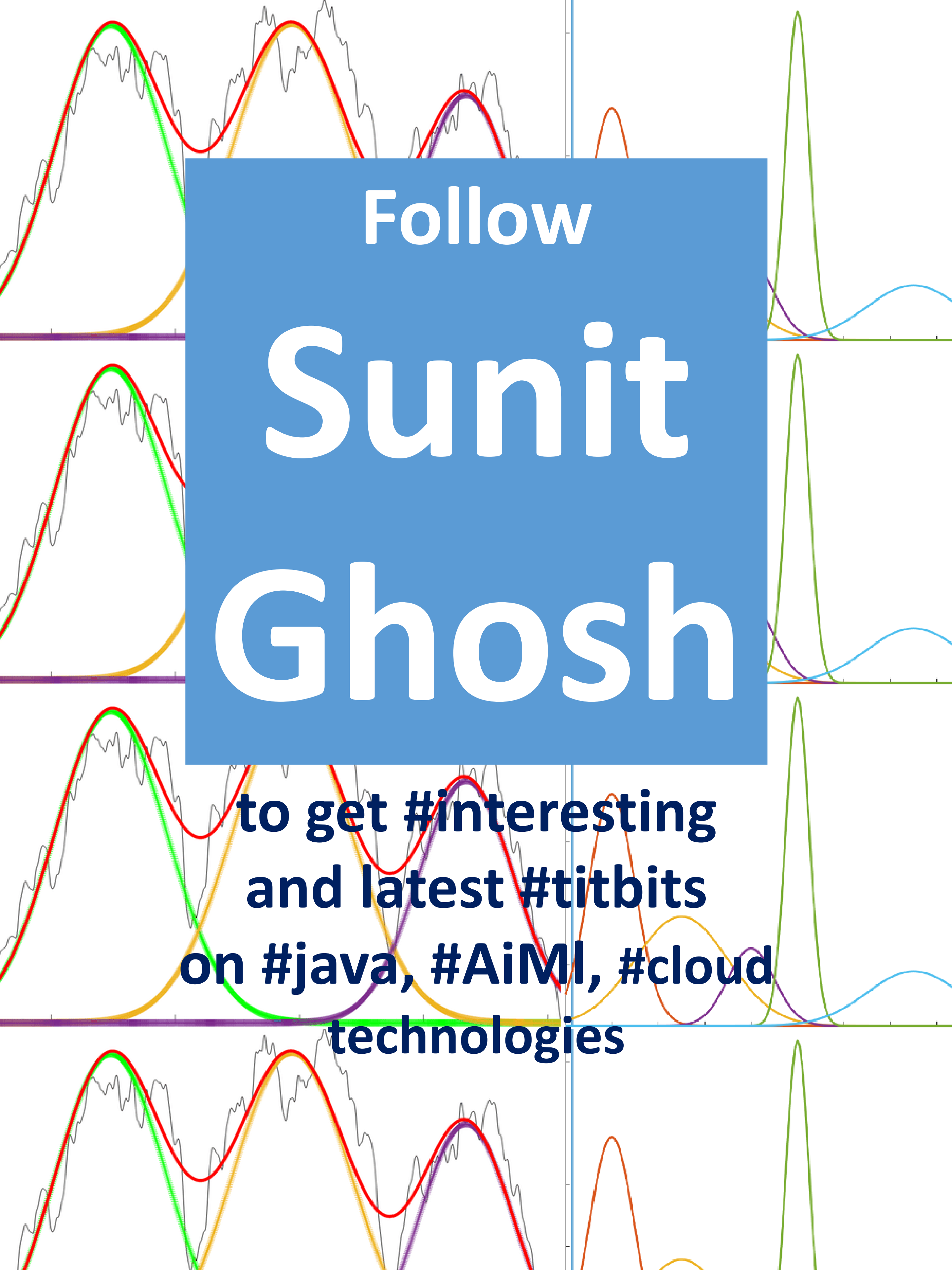
Real-world applications

In segmentation, GMMs are used to model the distribution of pixel intensities in an image and separate the foreground from the background

In speech recognition systems to model the distribution of speech features, such as mel-frequency cepstral coefficients (MFCCs), and classify them into phonemes

In protein structure prediction, GMMs can be used to model the distribution of atom positions and identify the most likely protein structure

Sunit



Follow

Sunit Ghosh

to get #interesting and latest #titbits on #java, #AiMl, #cloud technologies